Title

Considerations for the implementation of machine learning into acute care settings.

Journal
British Medical Bulletin, 141(1)

ISSN
0007-1420

Authors
Bishara, Andrew
Maze, Elijah H
Maze, Mervyn

Publication Date
2022-03-21

DOI
10.1093/bmb/ldac001

British Medical Bulletin, 2022, **141**:15–32
https://doi.org/10.1093/bmb/ldac001
Advance Access Publication Date: 28 January 2022

OXFORD

Invited Review

Considerations for the implementation of machine learning into acute care settings

Andrew Bishara[1,2], Elijah H. Maze[3], and Mervyn Maze[1,4,*]

[1]Department of Anesthesia and Perioperative Care, University of California San Francisco, 1001 Potrero Avenue San Francisco, CA 94110, USA, [2]Bakar Computational Health Sciences Institute, University of California San Francisco, 490 Illinois Street, San Francisco, CA 94143, USA, [3]Departments of Computer Science and Mathematics, University of Michigan, Bob and Betty Beyster Building, 2260 Hayward Street Ann Arbor, MI 48109, USA, and [4]Center for Cerebrovascular Research, Building 10, Zuckerberg San Francisco General, 1001 Potrero Avenue, San Francisco, CA 94110, USA

*Correspondence address. Department of Anesthesia and Perioperative Care, University of California, San Francisco & Zuckerberg San Francisco General, 1001 Potrero Avenue, San Francisco CA 94110-3519, USA. E-mail: mervyn.maze@UCSF.edu

Editorial Decision 1 January 2022; Accepted 1 January 2022

Abstract

Introduction: Management of patients in the acute care setting requires accurate diagnosis and rapid initiation of validated treatments; therefore, this setting is likely to be an environment in which cognitive augmentation of the clinician's provision of care with technology rooted in artificial intelligence, such as machine learning (ML), is likely to eventuate.

Sources of data: PubMed and Google Scholar with search terms that included ML, intensive/critical care unit, electronic health records (EHR), anesthesia information management systems and clinical decision support were the primary sources for this report.

Areas of agreement: Different categories of learning of large clinical datasets, often contained in EHRs, are used for training in ML. Supervised learning uses algorithm-based models, including support vector machines, to pair patients' attributes with an expected outcome. Unsupervised learning uses clustering algorithms to define to which disease grouping a patient's attributes most closely approximates. Reinforcement learning algorithms use ongoing environmental feedback to deterministically pursue likely patient outcome.

Areas of controversy: Application of ML can result in undesirable outcomes over concerns related to fairness, transparency, privacy and accountability. Whether these ML technologies irrevocably change the healthcare workforce remains unresolved.

Growing points: Well-resourced Learning Health Systems are likely to exploit ML technology to gain the fullest benefits for their patients. How these clinical advantages can be extended to patients in health systems that are neither well-endowed, nor have the necessary data gathering technologies, needs to be urgently addressed to avoid further disparities in healthcare.

Key words: acute care, machine learning, algorithms

Introduction

Ever since the Industrial Revolution, which introduced steam- and water-power, societies have largely embraced automation because of the benefits provided by removal of the drudgery of repetitive tasks, enhancement of efficiencies to reduce scarcities of products and services and improvements in the overall well-being of its inhabitants. As with subsequent technological revolutions involving electricity and computing, the Intelligence Revolution, exemplified by artificial intelligence (AI) is being heralded as a panacea for the well-being of the individual accruing from the additional time available for non-vocational pursuits and from the equitable distribution of financial rewards that automation can provide.[1]

At the personal level, benefit that results from improvement in one's health is a lofty goal that can be achieved through appropriately applied AI.[2] This optimistic outlook is predicated by several factors including the voluminous and heterogeneous medical data that is available at high speeds of access and analysis and generated through disparate sources such as claims data, intensive care surveillance, electronic healthcare documentation and medical device sensing and surveillance.[3] Because of the characteristics of these data the promise of personalized medical care is within reach through the blending of AI with research-generated population-level clinical evidence and the specific characteristics of the individual patient.[4] In this article, we describe how AI can be implemented into decision-making for acute care with the goals of reducing medical error (e.g. by reducing fatigue from attention-requiring tasks) and improving outcomes in a rapidly evolving and dynamic clinical setting.

As with each successive revolution in automation, safeguards are needed to curtail unwanted consequences that can disrupt the workplace. In its influential report on human-centric AI, the European Union (EU) stressed that the technology is not an end itself but a tool that has to serve the people by preserving the universal values of fairness, transparency, privacy and accountability; humaneness in the medical applications of AI has been stressed by others.[5] In addition to these necessary safeguards, another requirement for implementing successful and equitable applications of AI into the acute care setting will be the development of a sustainable economic model; this will also be considered in this report.

As this report is focused at the level of individual patients in acute care settings, we will not address the application and consequences of AI-based automation in non-clinical aspects of healthcare such as administration (e.g. scheduling appointments), finance (e.g. billing and collections) and operations (e.g. inventory and supply chain management). Because we chose to concentrate on the use of AI for a medical decision support system (MDSS) in the acute care setting, this report does not address other important AI applications such as image interpretation and remote monitoring, which may also enhance personalized care.

Sources of data for this review

PubMed and Google Scholar were the primary sources for this report; for searches of these databases, the terms used included: machine learning (ML), telemedicine, intensive care unit, randomized clinical trials, clinical decision support, anesthesia information management systems, database structure, electronic health records, causal inference reinforcement learning, acute kidney injury, ML versus physician and cardiac anesthesia.

Definitions

In this review, numerous terms are used, which require further definition. Although essential for the inexperienced practitioner, it is also necessary for those with expertise in this field because these ubiquitous terms may have different interpretations.

Artificial intelligence (AI)—a term first coined during a Dartmouth College summer conference in 1956—is a technique by which machines replicate the behavior and nature of humans.[6] The 'Turing Test' states that if a computer can convince a human that it is human, then the machine can 'think' and therefore has AI.[7] AI optimizes processes to operate autonomously and can result in complex outcomes that may extend beyond what was explicitly programmed. Currently, our understanding of true AI has evolved into a question of whether machines can achieve 'general intelligence'. In the medical setting, we will discuss the example of computers that 'learn' from real-world data that resides within electronic health records (EHR); ML, is the outstanding example of AI in the medical setting and is defined below.

Machine learning (ML)—a term first coined by Arthur Samuel[8]—is a form of AI in which a computer learns from historic data generating algorithms to apply rules to new data. As the robustness of ML depends on the volume of data, the abundance of accessible data has made it possible for these programs to be exceptionally powerful.

Neural network is a form of ML using a structure that mimicks the human brain and its networks of neurones; conceptually, the individual neurones within the input and output layers are assigned parameters that determine the level of effect propagated to the next neurone in the network. In the learning process, the output layer feedbacks on prior layers further refining the model, by modifying individual neuronal parameters, through a process known as backpropagation.[9]

Deep learning is a type of neural network that contains several layers between the input and output layers; these multiple, hidden, layers provide a 'deep' infrastructure improving the model's ability to learn complex patterns and abstractions that is best exemplified by image processing.[10] Deep learning has been criticized as 'black box' modeling that precludes interpretability because of the difficulty to understand and/or see the inner workings of the model.[11] Although tools have since been designed to aid with interpretability of deep learning models, these models cannot be equated to 'deep understanding'.

Medical decision support systems (MDSS) use computer tools, which do not necessarily require any level of ML or complex analytics, to help clinicians provide better care to patients. MDSS can include reminders to give antibiotics at a particular time in a surgical case and/or to document certain pieces of information in the chart, which in turn may remind the clinician to perform an additional exam or nudge a physician's thought process potentially leading to a more complete diagnosis of the patient's condition.

Personalized care/precision medicine uses MDSS and often ML and other AI techniques to provide treatment options tailored to a specific patient's condition or 'clinical signature' e.g. their unique set of diagnoses, medications, demographics and genetics, among other factors. It is within the realm of precision medical care that AI is likely to generate the biggest advances in the practice of medicine.

Expert systems were an early form of ML in which a series of multi-layered 'if-then' statements

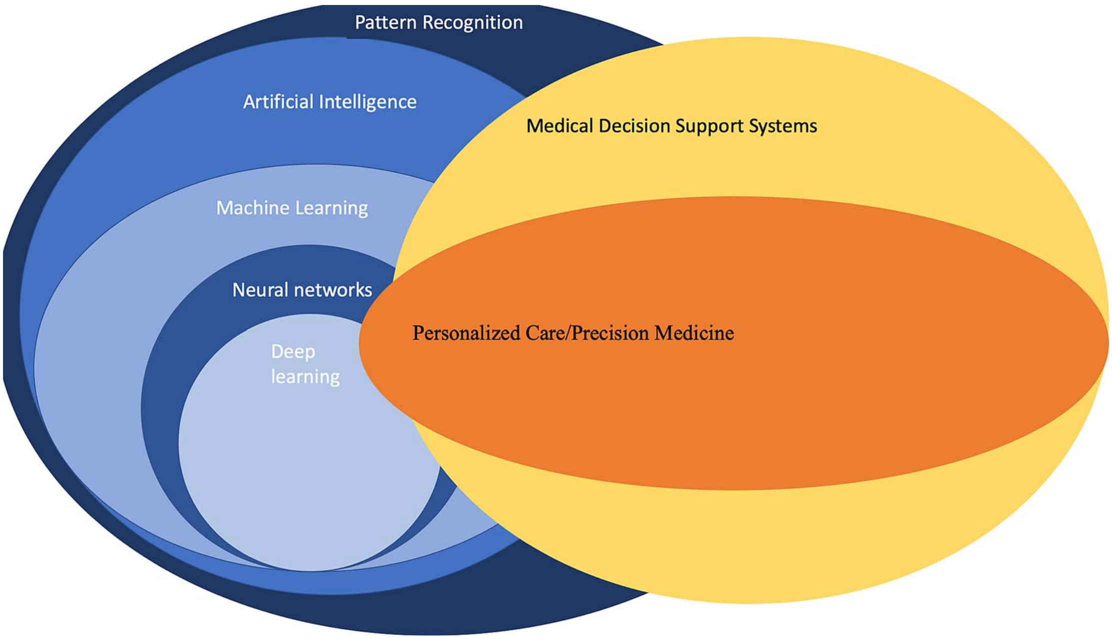

Fig. 1 In this Venn diagram the relationship of different forms of patient care digital technologies is displayed (modified from Vrana and Singh[68]).

determine the optimal course of action depending on a particular situation.[12] These systems rely on a set of Boolean and deterministic rules that are often directly programmed by experts in the field. This is no longer considered a form of intelligence because it neither infers meaning nor context.

The relationship of these terms to one another is depicted in Figure 1.

Logistics for development and implementation of ML

For the following section, the term ‘algorithm’ is a set of instructions that a computer follows to complete a task and a ‘model’ is the application of algorithms to data. Data scientists, model developers and clinical end-users, of the output of the model need to collaborate to successfully pair a validated prediction with an effective action. Foremost among the issues that determines the type of model to be developed is patient safety that include knowledge of the adverse risks to the patient if the model performs inaccurately. Decisions regarding the acceptable degree of accuracy will skew the model towards whether it is more desirable to have false positives or false negatives in the prediction of the association between the patient’s attributes and the anticipated outcome. Another important decision to be taken is whether the model is based upon defining association or causality.[13] As the full set of causal relationships is rarely known in acute care settings, most of the models that have been developed are association-based.

The family of algorithms that we call ML can be categorized depending on how they learn inference from data, namely, by supervised, unsupervised (Fig. 2) or reinforcement learning models.

Supervised learning—in its most general sense—is concerned with predicting outcomes for new pieces of data. Each piece of data upon which supervised learning algorithms are trained, or ‘learned’ from, consists of a list of attributes, or features, and an outcome that is referred to as a target or label. When this algorithm is trained on a multitude of these features and their corresponding labels, the model learns which features most closely correspond to each label via various mathematical algorithms and statistical models, generally, with a basis in linear

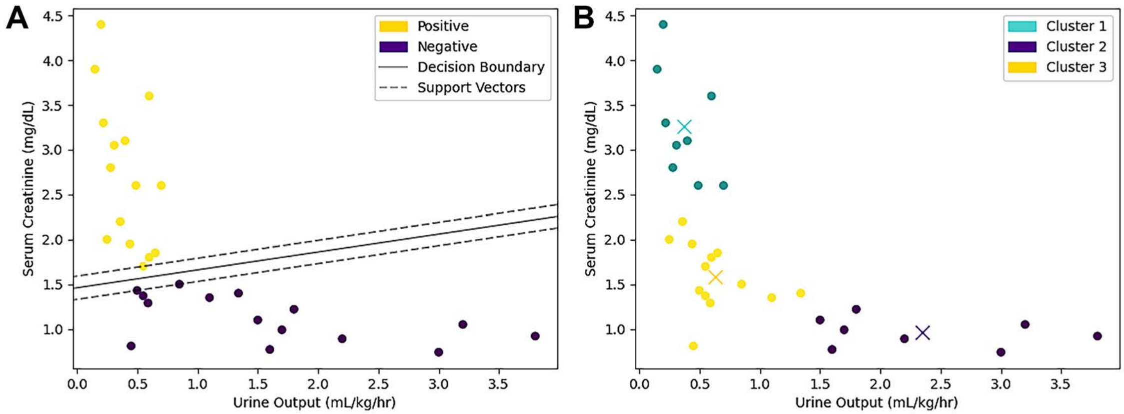

Fig. 2 A supervised learning model versus an unsupervised learning model with the same dataset of 30 patients: (**A**) Supervised Learning Model: This model uses a Support Vector Machine (SVM) that has a decision boundary to predict whether Acute Kidney Injury (AKI) will develop in any given patient with known features. In this theoretical example, a pair of features, namely, urine output and serum creatinine from 30 patients with a binary outcome (yellow for those that developed and blue for those that did not develop AKI) was used to train the SVM. From this training set, the SVM calculates the decision boundary for the development of AKI from the midpoint of the support vectors. In this situation, we have calculated the decision boundary to be the vector, which we refer to as θ. The trained SVM determines whether a new patient with a known pair of features (which we refer to as vector A) will develop AKI by calculating the dot product, $\theta \cdot A$ yielding a binary outcome depending on its sign (positive sign corresponding to a positive prediction, and *vice versa*). The complexity of the calculations that a SVM makes can be better appreciated when considering that the typical training set usually comprises many more than 30 patients (often in the thousands) with many paired features (likely hundreds rather than the two offered here), with the same result: a hyperplane separating vectors that will predict positive from those that will predict negative. (**B**) Unsupervised Learning Model: This model uses algorithms, and K-means, that implement pattern recognition. In this theoretical example the same paired features of urine output and serum creatinine from the 30 patients are separated into clusters using a clustering algorithm (the 3-means algorithm is offered here, with each 'mean', denoted in the graph by an 'X', representing the center of a cluster ['centroid']). In lieu of considering the outcome for each of the features as is done with supervised learning in the Supervised Learning Model, this clustering algorithm determines which patients are most like one another. Whereas the SVM provides a sharp prediction derived from features of a new patient (without considering how 'close' it is to predicting the opposite), clustering achieves a nuanced view of a patient's profile. Again, the model's complexity increases exponentially with the number of paired features.

algebra and probability theory. After this model has been trained, the user can then apply a novel set of data with known attributes, but an unknown outcome, into the supervised learning model, from which the outcome for each piece of data is predicted; possible deviations from the outcome or target may be quantified by an error function where feasible. Although these supervised learning processes may approximate a provider's skills, the computer can uncover novel relationships not readily apparent to physicians.

For the purposes of developing the algorithm for supervised learning we will use an example that predicts an outcome; in acute care settings the outcome can be a patient's diagnosis that will then engender a response such as initiating a known treatment for that diagnosis. If we wish to know for male patient 'X', whether certain of his attributes will hasten the development of acute kidney injury (AKI), we can train a supervised learning algorithm on past patients with these attributes and whether they developed AKI. After training, we can input patient X's attributes into the model, and it will predict whether patient X will develop AKI based on trends from all previous patients. Depending on which supervised model we use, it may determine which attributes most strongly correspond to developing AKI using a supervised learning algorithm. Here, we use an example of a seminal supervised learning algorithm, referred to as the support vector machine (SVM), which builds a relatively accurate classification or regression model.

The SVM algorithm constructs an optimal hyperplane to maximally separate two classes.[14] Certain features or attributes have been selected from the training data which, for the purposes of this example, we have narrowed to five (A–E; e.g. serum creatinine, blood pressure, fluid intake, urine output and exposure to nephrotoxins) for the outcome of development of AKI; for parameter fitting, the algorithm envisions these sets of attributes as mathematical vectors in a vector space with defined boundaries also referred to as 'separating hyperplanes'.[15] The goal of the algorithm is to learn a parameter hyperplane, θ, which separates the vectors that developed AKI from those that did not. In the case of a Linear SVM, when we are interested at making a prediction for a new patient, with attributes $A = [A', B', C', D', E']$, we simply take the product A dot θ. If this dot product is positive, A projects to the positive side of θ, so we predict the likelihood for the development of AKI. Likewise, the most positive elements of θ correspond to the features being most likely to develop AKI, as these will affect the result of the dot product most egregiously. An example of a two-feature SVM is depicted (Fig. 2). For some datasets, this linear approach may not be sufficient; in that case the SVM can adopt a non-linear approach in which case kernel functions are used to map the data to a higher dimensional space, where instead of taking a dot product, the model uses kernel algebra to determine where the input vector lies in relation to the separating hyperplane.[16] An SVM utilizing a non-linear kernel can lead to higher accuracy metrics; however, this occurs at the expense of losing interpretability, due to a higher level of abstraction.

After developing a model with minimal training-error, the generalisability of the model must be established by determining its predictive performance on a comprehensive set of informative features from previously unseen individuals to confirm low test-error. Many methods exist for such affirmation, notably cross-validation, which 'folds' the known data into test data and training data, so we can test the model on existing data efficiently. Among the reasons why the performance of a model deteriorates between training and testing include 'overfitting' in which a small set of training data has selected a very large number of attributes that 'memorises' rather than identifies the features and structures (including the noise) of the training set; as memorisation of noise within the training set is not generalisable it performs worse on the unseen test data. Although cross-validation is effective in identifying such errors, there are mathematical toolkits available to ensure regularisation in which the number of selected features and its vector space are curtailed resulting in less variance between the test and training sets, often at the expense of increasing training error. There is no perfect solution to this dilemma; ML is best when training is performed on a very large set of informative features, a circumstance that may be lacking for the complex tasks of clinical predictions based upon accessible data. Because of this constraint it is quite useful to have domain-specific clinician expertise to 'guide' the learning process.

The performance of these binary classification supervised learning models can be assessed by well-defined metrics mostly describing either sensitivity or specificity. Sensitivity (also referred to as true positive rate, probability of detection or recall) reflects the ability to identify the true positive cases; a highly sensitive model can reliably rule out a disease when its result is negative. Specificity (also known as the true negative rate) quantifies the portion of actual negatives that are correctly identified as such. The trade-off between sensitivity and specificity can be visualized by the receiver operating characteristic (ROC) curve, which is often summarized as the area under the ROC although this single parameter may not expose subtleties introduced by events such as imbalanced datasets where the negative and positive outcomes are not equally distributed.[17]

Unsupervised learning algorithms, on the other hand, do not serve as prediction tools (i.e. whether an outcome or label will be achieved), but rather they serve to identify correlations and internal structure ('clusters') in training data, which is often heterogeneous as is the case with most diseases. For unsupervised learning a clustering algorithm, such as the

K-Means Clustering algorithm, the model is trained solely on the attributes of each piece of data (i.e. features), and not the corresponding outcomes (i.e. labels).[18] During training, this type of algorithm identifies structures and patterns within the data; when new data are inputted into the model, it identifies the previous data to which the features are most similar. The model that is generated represents the training attributes in vectors, exactly like the SVM, but, instead of learning a boundary vector, it learns vectors that serve as the center ('centroid'), or mean, of each cluster with the number of clusters denoted as 'k'. Each of the training samples is assigned to one of the clusters according to a predefined distance metric from the center of the cluster. K-Means Clustering can be optimised by reducing the number of dimensions in which the data are visualized using techniques such as Principal Component Analysis. In the context of medical decision-making, if one were to input a patient's data into a clustering algorithm, the model would find the mean vector to which the patient's attributes are closest. This identifies to which previous patients our current patient is holistically most similar, so medical decisions can be determined more accurately based on the data of past patients in the same cluster. When K-Means Clustering was used on >11 000 intensive care admissions, membership of a cluster accurately predicted mortality, length of stay, requirement for mechanical ventilation and the use of inotropes.[19] An example of clustering on the basis of a pair of clinical features is provided (Fig. 2B). Unsupervised learning can also be used to identify dependencies between different data attributes. These models, most notably Bayesian Networks, are probabilistic models that learn how one particular attribute causes or affects other attributes. For example, it can learn what the contributing factors are to developing a particular attribute, or possibly once that attribute occurs, which effects are likely to follow. A Bayesian network simply learns a directed graph, with each feature as a node, where dependencies can be identified by the graph's edges and its direction. Bayesian processes are frequently used in medicine, most notably diagnosis of disease. ML networks are an augmentation of those practices, as it offers a comprehensive view of which features of a patient's profile are most likely to have an impact on others.

As indicated earlier, a blend of unsupervised and supervised learning may be required. A good example would be first to use unsupervised learning to select robust features that is then incorporated into the risk-prediction performed by supervised learning. In this form of Deep Learning for complex acute care, the blending may reclassify patients into more homogeneous groups in which the pathophysiology, and response to therapy, may be shared.

Reinforcement learning is more akin to human learning than either supervised or unsupervised learning, as the model interacts with the environment and continues to train itself during implementation and prediction. As it receives feedback from the environment there are a large number of possible actions that the model can make at each instance; in some cases, the training space is virtually infinite as actions that it takes can change the environment. In other cases, when reinforcement learning starts, every action is random and hence exploratory, as it registers the feedback from the system; when sufficient feedback has been gathered the actions change from random to deterministic, which is referred to as exploitation. In this particular format of reinforcement learning there is constant trade-off between exploration and exploitation; although exploitation accrues short-term rewards, it precludes greater long-term rewards that are currently not known. These models have shown impressive results in the realm of recommending optimal treatments for sepsis in the intensive care unit (ICU).[20] Within acute care settings, reinforcement learning comes to the fore, enabled by the short latency between treatment and response to treatment that facilitates iterative processing. For example, vasopressor dosing is constantly revised to reach an appropriate infusion rate for a particular patient. The optimal dosing can initially be estimated based on weight, cardiac function, age, volume status and other factors, but cannot be known until a dose is attempted and the blood pressure and heart rate response are observed. Although a rudimentary example of this is target-controlled infusions, reinforcement learning allows for more comprehensive control that can considers optimal dosing of many infusions in the context of an individual patient's medications and past medical history.

We imagine that reinforcement learning will play a large role in the future integration of AI into the practice of medicine as these models will allow the model to interact with the medical team and to learn tendencies of clinical providers, as well as new patient cases, as they are used. These models can also recommend specific treatments or combinations of treatments and may provide creative treatment strategies not otherwise considered.

Why precision medicine facilitated by ML is desirable in acute care

In this section, we posit that through the application of *in silico* biological intelligence tools, medical professionals can provide personalised/precision care with a reduction in medical errors.

Acute care lends itself to improvement through ML because of its rapidly and constantly growing volume of data which, for the most part, captures a large component of the decision-making information used by physicians. Data streams from intensive care monitors provide the best example of a data source that reflects the continuous measurement of a patient's health, albeit for a short duration. For these reasons, telemedicine has flourished within acute care settings with providers able to analyse the digitised data remotely.[21] As the quantity of data increases, ML improves its ability to make fast accurate decisions through a network of computations that iteratively train the models. The model can consider the data of the individual patient being cared for and compare that patient's condition with all similar patients with analogous clinical courses (meaning they have similar pathology, past medical history and treatments) to make precise and accurate predictions and treatment recommendations.[15] An excellent example of the utility of ML for risk prediction has been the early identification of ICU patients that are likely to develop sepsis including the use of physiomarkers in the pediatric setting.[22] A meta-analysis demonstrated that ML-based tools are superior to other methods of scoring risk for developing sepsis[23]; however, between-study heterogeneity may thwart the potential advantage.[24] A non-comprehensive set of examples of successful ML-based application in the acute care setting is tabulated (Table 1).[25–37]

These ML tools, when visualized in an easy-to-understand manner, can provide value to critical care physicians who are often inundated with an overwhelming quantity of data from each patient. It is anticipated that in a similar manner to the airline industry, computers will improve safety in this clinical setting.[38]

As there have been instances where ML has performed better than clinicians (AKI in the ICU[39]; AKI prediction after cardiac surgery[40]), we can anticipate situations for which ML can provide predictions of outcomes and recommended treatments to reduce medical errors and to also increase the capabilities of modern healthcare to improve likelihood of recovery from acute illness. In fact, there have been clinical studies showing that ML-based early-warning systems in the operating room can reduce total number of minutes of hypotension during anesthesia.[25] As these technologies develop, we will be able to predict which patients will deteriorate earlier and be able to provide more specific and tailored treatments for those patients. We will also be able to provide more complex rules that allow for more nuanced treatment guidelines and circumstances, which allow us to capture more subpopulations of patients and more 'edge cases' for whom specific treatments might be harmful or beneficial. As the technologies become better at explaining why predictions and recommendations were made, physician experts and AI will more effectively complement one another to provide better treatment than either could provide alone. In a recent study performed in the ICU setting, personalised medical care was provided through expert-augmented ML in which, iteratively, AI learns from the input and the tendencies of clinicians to efficiently improve patient care.[41]

Table 1 Examples of Machine Learning in Acute Care

Author	Year	Study type (*n*)	Aim	Conclusion
Wijnberge *et al.*[25]	2020	RCT of ML-based EWS vs standard care (68)	Limiting intra-op hypotension	Reduction in hypotension
Xue *et al.*[26]	2021	Retrospective analysis (111 888)	Predicting postop delirium	Predictive
Rehm *et al.*[27]	2018	Prospective observational (35)	Predicting patient-ventilator asynchrony	Predictive
Zhang *et al.*[28]	2021	Prospective observational (459)	Predicting agitation on ventilator	Predictive
Hsieh *et al.*[29]	2018	Retrospective analysis (3602)	Predicting successful extubation from MV	Predictive
Hur *et al.*[30]	2021	Retrospective analysis (12 409)	Predicting ICU delirium	Predictive
Le *et al.*[31]	2020	Retrospective analysis (9919)	Predicting development of ARDS	Predictive
Sinha *et al.*[32]	2020	Secondary analysis of RCT patients (2767)	Diagnosing ARDS phenotype	Diagnostic of 2 phenotypes
Le *et al.*[33]	2019	Retrospective analysis (101)	Predicting pediatric sepsis	Predictive
Nema *et al.*[34]	2018	Retrospective analysis (69 000)	Predicting adult sepsis	Predictive
Seymour *et al.*[35]	2019	Retrospective analysis (20 189)	Diagnosing sepsis phenotype	Diagnostic of 4 phenotypes
Zhang *et al.*[36]	2019	Retrospective analysis (6682)	Predicting UO in response to fluid bolus	Predictive
Lee *et al.*[37]	2018	Retrospective analysis (2010)	Predicting postop acute kidney injury	Predictive

ARDS = Acute Respiratory Distress Syndrome; ICU = Intensive Care Unit; MV = mechanical ventilation; UO = Urine Output

A theoretical application of ML in the acute care environment is provided (Fig. 3); this example is notable for its dependency on the EHR. However, the difficulty with the EHR as the source of information is that it stores a mixture of structured and unstructured data, often inconsistently formatted, full of variables with varying quantities of missing data and at times does not actually reflect what occurred clinically. A clinical AI system would likely need to learn from more than just the findings within the EHR to care for patients.

A roadmap for the development, implementation and assessment of an ML-based solution to management of patients in clinical settings, including acute care, has been formulated.[42] Beginning with elucidation of a clinical problem for which an improved outcome is sought, the process involves early engagement of stakeholders, and a clear message how adoption of the technology will improve workflow and patient outcomes. Most importantly, the plan needs to define the milestones, metrics and outcomes that determine whether an implementation is successful using, amongst others, the framework provided by the International Medical Device Regulators Forum. The AI tools that are developed for an acute care setting are high risk requiring a premium on their assessment of safety and effectiveness.

Finally, as with all deployments in a Learning Health System, an after-action assessment will be instrumental to inform further implementations.

Obstacles to implementing ML in acute care settings

Safety/Accountability/Liability

Building precision care systems and integrating them into clinical care are not the same problem although they have often been discussed as such. The ability of a ML model to predict in new situations is assessed by that model's ability to accurately predict on a 'test' dataset. To ensure that ML models can predict well in real-world settings, the test set must

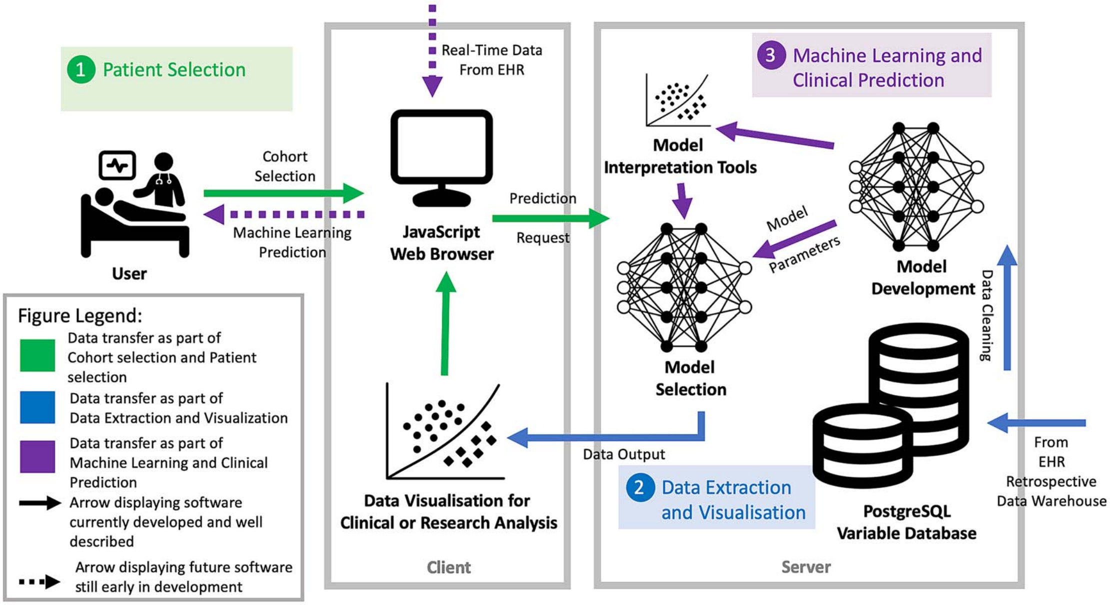

Fig. 3 Overview of a possible dataflow structure for a ML-based MDSS. The dataflow here is outlined in three phases: patient selection, data extraction and visualization, and ML and clinical prediction. Models are initially trained after extraction from the EHR and data cleaning as necessary. Model parameters are uploaded to the software infrastructure that integrates with the EHR and allows for real-time data updates for patients currently being cared for in an acute setting. A patient is specified by the user and the appropriate model is selected depending on the prediction required. Once data have been passed to the client, the user has the opportunity to visualize the prediction and reasons for the prediction on the dashboard.

accurately represent the real-world cases to which the model will be applied (see Fig. 1 in Cosgriff *et al.*[2]), which is a fundamental reason why a model can seem to perform well in development, but then provide inferior accuracy when actually deployed. This raises a serious safety problem that models need to be well-tested in restricted clinical settings before deployed to many more patients.

Typically, a medical error by a provider may result in injury to a single patient; a software problem in a ML-based MDSS can cast a much wider net and injure many more patients, in the same way that an inaccurate test or biomarker can incorrectly guide treatment. Furthermore, patients are likely to be more forgiving of a medical error produced by a human provider than through a software error. However, in litigious settings, even though a software error caused the problem to the patient, the provider should have relied on their own professional knowledge and judgment akin to the development of self-driving motor vehicles in which the human driver still has some control during its early evolution (e.g. see Fig. 5 in Topol[43]). Therefore, clinical implementation of an effective MDSS based upon an ML-derived algorithm may be resisted by a healthcare system because of the potential for unsustainable litigation costs caused by the false-positive and false negative cases or the incorrect estimations and predictions developed by fully-automated systems (*vide infra* 'AutoML'). To prevent this negative outcome, every effort needs to be expended to demonstrate the robustness of the ML-based models through rigourous training and testing under real-world conditions in the same settings in which it is to be clinically used. There are likely to be exceptions to following a decision promoted by the predictive analytics; under these circumstances the consultant physician needs to 'play their traditional role of patient advocate within the constraints set by society' and not to abrogate their responsibility to software.[44] Although the goal of the model may be to establish clinical endpoints (such as the likelihood of developing postoperative acute kidney failure), it is also important to understand the model's effects on patient-centered outcomes as well as the work-flow issues for practitioners.

It is certainly the case that overriding alerts may prove damaging in a litigious setting.[45] A provider may possibly be considered liable for clinical negligence if she/he rejected the output of the MDSS and there was subsequent harm to the patient. Contemporaneously, recording the reasons for the inaction, in conjunction with the patient's awareness of the situation (where applicable), may be the prudent route forward.

Interpreting ML findings to patients

When considering interpretability, it is well to remember that patients may be neither computer- nor biologically-literate. Therefore, it is necessary to explain, in layman's terms, the multi-level models that provide the foundation for the mathematical algorithm, the functional forms of the input data and the expected outcome or output. Furthermore, trust of the patient in the provider that uses opaque algorithms constructed from big datasets may be eroded when there are intellectual property issues involved in which privacy-justified secrecy is required.[46] In these circumstances, the performance characteristics of the models need to be continually evaluated including its sensitivity, its positive and negative predictive values and its ability to distinguish between positive and negative outcomes.[44]

Practitioners who use a ML-based MDSS must be able to interpret how they reached specific decisions (e.g. why the model predicts the likely development of acute kidney failure that requires an intervention to thwart this predicted outcome) and must be able to explain those decisions to any patient affected by them; a system for doing this can involve back-propagation (see Fig. 1 in Lauritsen *et al.*).[47] Failure to provide understandable explanations, places the practitioner in the unenviable position of having to personally vouch for the trustworthiness of an ML-based MDSS; this becomes particularly problematic when the recommendation provided by the MDSS conflicts with guidelines of medical practice. Such a scenario can be avoided if current guidelines are entered as 'prior knowledge' into the model. According to the EU's General Data Protection Regulation, the patient has a right to explanation of all decisions made by 'automated or artificially intelligent algorithmic systems;' the 'data controller' is legally bound to provide requesting patients with 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such data-processing for the subject.'.[48]

Privacy/Anonymity

We live in an age of Big Data; sites visited on the internet, places travelled and purchases made are documented in a vast database creating a network of comprehensive personal profiles. The application of Big Data for societal benefit is at a crossroads occasioned by high profile nefarious practices such as that used by Cambridge Analytica in which it sold to political campaigns psychological profiles of American voters acquired from Facebook data from millions of unknowing users.[49]

Data collection may be deemed equitable when viewed as a risk/benefit continuum in which the provision of health data, while infringing on privacy, results in advances in healthcare that improve the patient's life. Concerns related to privacy breaches can be considered as consequentialist, if the patient has come to some harm (from personal data being 'out there') or deontological, if there has been wrong doing (i.e. accessing data without explicit consent) without the patient suffering direct harm[50]; the latter is an example of loss of control over one's own data that has become an ethical flashpoint of late.

The lawful access to medical data differs remarkably between the US and the EU. In the US, the Heath Insurance Portability and Accountability Act (HIPAA) was enacted by Congress in 1996 at a time when most patients' health data were contained in analogue health records and it focused on preventing lapses by healthcare providers and health systems, the so-called 'custodians' of protected health information (PHI); importantly, under the Common Rule governing research, the regulations do permit the sharing of de-identified PHI in which 18 specific identifiers are removed. Now, HIPAA appears outdated[51] as the health data ecosystem has increased to include not only data on healthcare but data on health that is collected by many products and devices[52]; consider for example data from Google searches about a particular symptom or insurance coverage for serious diseases. Furthermore, with triangulation of other publically-available documents it may be possible to 're-identify' the specific patient.

Future governance arrangements for the stewardship of health/healthcare data are being considered and can extend from an externally mandated or 'broad consent' along the lines that are used in the establishment of biobanks (see Fig. 4 in Mayer-Schonberger *et al.*[53]) to one that maximizes patient autonomy and essentially requires the patient to approve the use of each and any data.[54] It is expected that these governance arrangements will be unfeasible for either patients (through 'underprotection') or the big data community (through 'overprotection') and that middle ground, along the lines of the Independent Review Panels established for handling requests to share clinical trial data, may suffice.[55]

Ethics/Fairness/Equity

The basic biomedical ethical principles of respect for patient autonomy, non-maleficence, beneficence and justice/equitability should obtain for ML-based MDSS. These tenets are especially important to sustain because the various stakeholders may have conflicting priorities including reputation/reimbursement (health care institutions), workflow/income (providers) and functional abilities/quality of life (patients).

Patient autonomy should allow the patient to 'opt-out' as is the case in the informed consent processes for a clinical research setting. However, it can be argued that when a clinical decision is arrived at from a model-based algorithm it is in the domain of quality improvement, an activity that does not typically require the patient's consent *a priori*. Nevertheless, to build trust and transparency in new technology serious consideration should be given to whether consent is required.[56]

To be generalisable, models must be built on data that is representative of the whole population; governance structures for the collection of the data should include representation by groups that have been particularly disadvantaged by disparities in healthcare. Diversity in terms of gender, racial or ethnic origin, religion or belief, disability and age, should be ensured at every stage of data acquisition and model development. As a corollary, those whose data contributed to the model should proportionately enjoy its benefits as a matter of fairness. How this fairness doctrine would work in practice is difficult to envisage; it is conceivable that datasets used to build the model may have been obtained from Public Health and Veterans Administration Hospitals but the private vendors may price its models to make them unaffordable to these underfunded care facilities. Furthermore, the computational power required may be prohibitively expensive to use especially for models using kernel algebra for the SVM as a higher dimensional kernel greatly increases the number of computations required. If there are no mechanisms to resource the use of highly successful applications in hospitals that disproportionately treat patients from underserved communities, health disparities will widen.

Training data that have been acquired from sources biased by a particular gender, race, age or sexual orientation run the risk of inaccurately generalising to non-representative populations. This form of bias has plagued the application of AI to activities such as criminal justice sentencing and hiring.[57] Inequities of this type can be prevented by having diversity in the AI teams as well as in the end-users.

Regarding non-maleficence, any unintended harm should be avoided. It is conceivable that institutions may use the models off-label; e.g. the analysis may be used to 'identify high-risk, high-cost patients in order to exclude them from its health care system'.[56]

Data access and availability for ML and its generalisability

The quality of data depends on its accuracy, validity, completeness and availability; in developing AI solutions, choosing high quality is more important than which learning models to use.[58] Health data in the EHR is fragmented and exists in multiple systems and formats that may not all be accessible to the tools of AI. The difficulty with the EHR as the source of information is that it stores a mixture of structured and unstructured data, often inconsistently formatted, full of variables with varying quantities of missing data and at times does not actually reflect what occurred clinically. A clinical AI system would likely need to learn from more than just the findings within the EHR. Lessons need to be learned from the well-intentioned introduction of EHR which, at times, converted providers into data-entry clerks and was associated both with physician burnout and patient dissatisfaction.[59] The workforce, which had difficulty adapting to the introduction of EHR, may be sceptical about the subsequent integration of AI into EHRs.[60]

Solving the accessibility conundrum to healthcare data will increase the cost of collecting the necessary data and precludes all but the best resourced from developing effective health care AI. The ML communities have already worked on producing decentralised analytical solutions to bypass this bottleneck including the use of Common Data Models in which data from many sources are aggregated.[61]

Also, of importance is the lack of generalisability of many of these models. Although a model might perform well in one environment or hospital, moving to a new hospital might provide very different variable inputs and/or practice tendencies, referred to as dataset shifts, that can affect performance of the model and the safety of clinical decisions.[62] A well-designed model should consider many of these potential variations and ideally be trained on data from each of the places where it will be deployed, but this is not always possible. Furthermore, there are other potential issues with workflow integration that might occur in areas of new deployment, and some have argued that true generalisability is unachievable and that we nearly always need to retest 'off the shelf' models in new settings.[63]

Regulatory approval

ML-based software designed for the purposes of aiding diagnosis and treatment decisions are defined as medical devices and in the US and the EU these require formal approval from the regulatory agencies; to date, more devices have been approved by the EU than the US.[64] It is hoped that a common transatlantic regulatory net will be applied to this consequential emerging technology along the lines of harmonisation of drug approval by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA).

Economic considerations

A robust business model is needed to establish who controls the device and/or software and how others can access it. Building the device/software requires considerable resources including hardware, software engineers, data scientists, together with clinical data and the manner whereby the developers capture return on its investment is receiving careful scrutiny following the adverse consumer experience with other digital assistants (*vide supra* Privacy/Anonymity). Economists have commented that 'unless the markets for innovation are fully contestable, the surplus earned by innovators will exceed the costs of innovation'.[65]

As it seems unlikely that clinical providers will have the expertise in ML sufficient to build, implement and integrate the ML device within the clinical workflow, it may require the institution to buy and/or rent off-the-shelf devices. However, bought or rented devices will require customization and calibration with local datasets and workflows, which

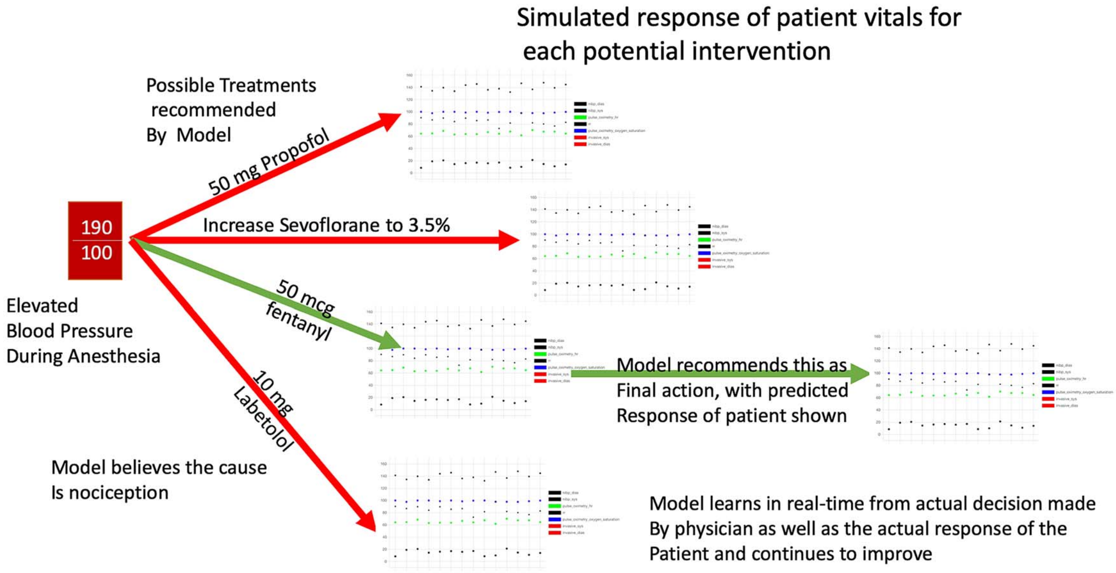

Fig. 4 Remedying a model to provide appropriate treatment for the potential diagnosis. The model identifies a situation that requires remediation, provides possible diagnoses for the problem, suggests possible treatments for each diagnosis, recommends the optimal treatment and foresees the expected patient response to the treatment chosen. 'Smart' trial design will be necessary for implementation of such future research in order to introduce the software in safe and constantly monitored ways.

will be costly and time-consuming.[63] It is possible that the vendor will request access to the local datasets to facilitate the customisation, which can introduce concerns about Privacy/Anonymity.

Another economically challenging aspect is the change in the workforce that may be provoked by ML-derived automation. As AI approximates human general intelligence, human labor may become obsolete. The Brookings Institute predicts that a third of all the tasks performed by healthcare practitioners may be automated. We anticipate that ML will develop intelligence-assisting ('IA') tools that support, but not necessarily replace, critical-thinking skills; as such we do not envisage wholesale reduction of the healthcare workforce in acute care settings in the guise of improving efficiency. Rather, we expect that the automation that ensues from AI may enable the workforce to do more good for more people especially in regions that have deficient access to specialty clinical care.[66] Nevertheless, the changes in the workforce could be profound including how staff work in a hospital or health system, the new skills and competencies required in the digital workplace, new AI positions, deployment of existing employees into other realms within the organization and how the overall workplace culture embraces innovation and technology (See Fig. 1B in Cosgriff *et al.*[2]). What was envisaged by an Institute of Medicine report on a 'Learning Health System' was a collaboration between the ML and acute healthcare communities in the pursuit of methods, protocols, guidelines and data analysis pipelines that explicitly take into consideration societal issues.[67]

Solving the accessibility conundrum to these data will increase the cost of collecting the necessary data and precludes all but the best resourced from developing effective health care AI. Solutions should be sought to lower the cost of AI technologies in order that disparities in the care delivered does not occur in resource-constrained environments, which are especially likely to exist in rural areas.

A roadmap for future research

As physicians discharge their social responsibility to seek better ways to improve the care of their patients, they need to exploit the opportunity provided by AI. An example of future research is illustrated (Fig. 4). However, more is required to achieve the goal of methodically and safely integrating AI into acute care to provide patients with more creative treatment strategies while avoiding adverse events. The clinical and cost effectiveness of an AI solution for acute care settings will require measurements of its utility, feasibility, implementation costs, clinical uptake and its maintenance of functionality over time.

Causality modeling

As mentioned earlier, models tend to be based upon associations as the causal relationships for most acute care conditions are not yet known. In the setting of considering more in-depth representations of the connection between inputs and outcomes (e.g. physiological mechanisms) we can imagine that with sufficient 'background knowledge' and additional explanations of the underlying connections between inputs, the model could build diagrams that represent the causal relationships between the variables considered. In the future, it will be desirable to develop models that can actually describe the causal mechanisms why a particular treatment will be appropriate for a particular patient with specific disease processes. As a corollary to such developments, it is conceivable that causality models could even be used to discover novel mechanisms and drugs for disease treatment.

AutoML: As the name implies, it is a system in which ML problems are solved through automation; in its ultimate rendition AutoML seeks to both build an entire ML pipeline and to optimise it automatically within the hardware resources available. As efficient as this promises to be, AutoML introduces further complexity when interpreting ML findings for patients. It seems unlikely that acute care providers will abrogate their accountability to AutoML; rather, these clinicians, will use ML solutions as cognitive augmentation in much the same way that they currently consult with other medical experts to redress their uncertainties concerning the patient's condition.[68]

Conclusions

In the dynamic and time-constrained setting of acute care, decisions are arrived at, and actions taken, before the extent of the patient's condition can be fully understood. It is in this setting that ML-based MDSS are likely to come to fore throughout the clinical care process including prevention, early detection, risk/benefit identification, diagnosis, prognosis and personalised treatment; however, the potential for beneficence may not be realised unless the potential obstacles are considered and mitigated. It is hoped that technological breakthroughs will be successfully combined with bioethical considerations so that patient safety and outcomes can be improved in the acute care settings. With the interaction of acute care clinicians and their patients with bioethicists, data scientists and lawyers, the collaboration can put in place governance structures that not only preclude new inequities but address ones that currently exist in acute care.

Data availability

No new data were generated or analysed in support of this review.

Conflict of interest statement

The authors have no potential conflicts of interest.

References

1. Zhang S, Srinivasan K, Singh PV, Mehta N. *AI Can Help Address Inequity–If Companies Earn Users Trust*. Harvard Business Publishing, Brighton, MA, USA: Harvard Business Review, September 17, 2021

2. Cosgriff CV, Stone DJ, Weissman G, et al. The clinical artificial intelligence department: a prerequisite for success. *BMJ Health Care Inform* 2020;27:e100183.
3. Matheny M, Israni ST, Ahmed M, et al. *Artificial Intelligence in Health Care; The Hope, the Hype, the Promise, the Peril*. Washington D.C., USA: National Academies Press, 2019
4. Executive Office of the President. *Big Data: Seizing Opportunities, Preserving Values* https://bigdatawg.nist.gov/pdf/big_data_privacy_report_may_1_2014.pdf; 19 June 2021, date last accessed 2014.
5. Celi LA, Fine B, Stone DJ. An awakening in medicine: the partnership of humanity and intelligent machines. *The Lancet Digital Health* 2019;1:e255–7.
6. Moor J. The Dartmouth College artificial intelligence conference: the next fifty years. *AI Magazine* 2006;27:87–91.
7. Turing AM. Computing machinery and intelligence. *Mind* 1950;236:433–60.
8. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 1959;3:210–29.
9. Lillicrap TP, Santoro A, Marris L, et al. Backpropagation and the brain. *Nat Rev Neurosci* 2020;21:335–46.
10. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527–54.
11. Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med* 2018;6:216–27.
12. Giarratano J, Riley G. *Expert Systems: Principles and Programming*. Boston, MA, USA: PWS Publishing Company, 1994
13. Breiman L. Statistical modeling: the two cultures. *Stat Sci* 2001;16:199–231.
14. Vapnik VN, Lerner AY. Pattern recognition using generalized portrait methods. *Autom Remote Control* 1963;24:774–80.
15. Ohu I, Benny PK, Rodrigues S, Carlson JN. Applications of machine learning in acute care research. *J Am Coll Emerg Phys Open* 2020;1:766–72.
16. Vapnik VN. *The Nature of Statistical Learning Theory*, 2nd edn. New York: Springer, 1995
17. McClish DK. Analyzing a portion of the ROC curve. *Med Decis Making* 1989;9:190–5.
18. Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. *J Royal Stat Soc Ser C (Appl Stat)* 1979;28:100–8.
19. Williams JB, Ghosh D, Wetzel RC. Applying machine learning to pediatric critical care data. *Pediatr Crit Care Med* 2018;19:599–608.
20. Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018;24:1716–20.
21. Breslow MJ, Rosenfeld BA, Doerfler M, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: an alternative paradigm for intensivist staffing. *Crit Care Med* 2004;32:31–8.
22. Kamaleswaran R, Akbilgic O, Hallman MA, et al. Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the PICU. *Pediatr Crit Care Med* 2018;19:e495–503.
23. Islam MM, Nasrin T, Walther BA, et al. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed* 2019;170:1–9.
24. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020;46:383–400.
25. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning sustem for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* 2020;323:1052–60.
26. Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open* 2021;4:e212240.
27. Rehm GB, Han J, Kuhn BT, et al. Creation of a robust and generalizable machine learning classifier for patient ventilator asynchrony. *Methods Inf Med* 2018;57:208–19.
28. Zhang Z, Liu J, Xi J, et al. Derivation and validation of an ensemble model for the prediction of agitation in mechanically ventilated patients maintained under light sedation. *Crit Care Med* 2021;49:e279–90.
29. Hsieh M-H, Hsieh M-J, Chen C-M, et al. An artificial neural network model for predicting successful extubation in intensive care units. *J Clin Med* 2018;7:240.
30. Hur S, Ko R-E, Yoo J, et al. A machine learning–based algorithm for the prediction of intensive care unit delirium (PRIDE): retrospective study. *JMIR Med Inform* 2021;9:e23401.
31. Le S, Pellegrini E, Green-Saxena A, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). *J Crit Care* 2020;60:96–102.
32. Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress

syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med* 2020;202:996–1004 PMID: 32551817.

33. Le S, Hoffman J, Barton C, et al. Pediatric severe sepsis prediction using machine learning. *Front Pediatr* 2019;7:413.
34. Nemati S, Holder A, Razmi F, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018;46:547–53.
35. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019;321:2003–17.
36. Zhang Z, Ho KW, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care* 2019;23:112.
37. Lee H-C, Yoon H-K, Nam K, et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. *J Clin Med* 2018;7:322.
38. Zenati MA, Kennedy-Metz L, Dias RD. Cognitive engineering to improve patient safety and outcomes in cardiothoracic surgery. *Semin Thorac Cardiovasc Surg* 2020;32:1–7.
39. Flechet M, Falini S, Bonetti C, et al. Machine learning versus physicians' predictiom of acute kidney injury in critically ill adults: a prospective evaluation of the AKIpredictor. *Crit Care* 2019;23:282–92 PMID: 31420056.
40. Rank N, Pfahringer J, Stamm C, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *NPJ Digit Med* 2020;3:139–50.
41. Gennatas E, Friedman JH, Ungar LH, et al. Expert-augmented machine learning. *Proc Natl Acad Sci U S A* 2020;117:4571–7.
42. Fihn SD, Saria S, Mendonça E, et al. Deploying AI in clinical settings. In: Matheny M, Israni ST, Ahmed M, Whicher D (eds.). *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*edited by. Washington, DC: National Academy of Medicine, 2019
43. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
44. Cohen GI, Amarasingham R, Shah A, et al. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff* 2014;33:1139–47.
45. McCoy AB, Thomas EJ, Krousel-Wood M, Sittig DF. Clinical decision support alert appropriateness:a review and proposal for improvement. *Oschner J* 2014;14:195–202.
46. Spector-Bagdady K. "The Google of Healthcare": enabling the privatization of genetic bio/databanking. *Ann Epidemiol* 2016;26:515–9.
47. Lauritsen SM, Kristense M, Olsen MV, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat Commun* 2020;11:3852.
48. General Data Protection Regulation (GDPR): http://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm 12 July 2021, date last accessed
49. Confessore N. *Cambridge Analytica and Facebook*. New York, NY, USA: New York Times, ISSN 0362-4331 April 4th, 2018
50. Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med* 2019;25:37–43.
51. Cohen IG, Mello MM. Big data, big tech, and protecting patient privacy. *JAMA* 2019;322:1141–2.
52. Greenberg MD, Ridgely MS, Hillestad RJ. Crossed wires: how yesterday's privacy rules might undercut tomorrow's nationwide health information network. *Health Aff* 2009;28:450–2.
53. Mayer-Schonberger VM, Ingelsson E. Big data and medicine: a big deal? *J Intern Med* 2018;283:418.
54. Kaye J, Whitley EA, Lund D, et al. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet* 2015;23:141–6.
55. Rockhold F, Nisen P, Freeman A. Data sharing at a crossroads. *N Engl J Med* 2016;375:1115–7.
56. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Crown Books Publishing, 2017
57. Garcia M. Racist in the machine: the disturbing implications of algorithmic bias. *World Policy J* 2016/17;33:111–7.
58. Kilkenny MF, Robinson KM. Data quality: "garbage in – garbage out". *Health Inf Manag* 2018;47:103–5.
59. Verghese, A. *How Tech Can Turn Doctors into Clerical Workers*. The New York Times, 2018 May 16. https://www.nytimes.com/interactive/2018/05/16/magazine/health-issuewhat-we-lose-with-data-driven-medicine.html). 7 December 2021, date last accessed
60. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016;165:753–60.

61. Rosenbloom ST, Carroll RJ, Warner JL, et al. Representing knowledge consistently across health systems. *Yearb Med Inform* 2017;26:139–47.
62. Subbaswamy A, Schulam P, Saria S. Preventing failures due to dataset shift: learning predictive models that transport. *Proce Mach Learn Res* 2019;89:3118–27.
63. Futoma J, Simons M, Panch T, et al. The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health* 2020;2:e489–92.
64. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *The Lancet Digital Health* 2021;3:e195–203.
65. Korinek A, Stiglitz JE. Artificial intelligence and its implications for income distribution and unemployment. In: Agrawal A, Gans J, Goldfarb A (eds.). *The Economics of Artificial Intelligence*. Chicago, IL, USA: Univ of Chicago Press, 2019, by
66. Mesko B, Hetenyi G, Gyorffy Z. Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Serv Res* 2018;18:545–8.
67. The Foundations for Continuous Improvement in Health and Health Care. *Digital Infrastructure for the learning health system*. Natl Acad Med, 2011
68. Vrana J, Singh R. NDE 4.0 – a design thinking perspective. *J Nondestr Eval* 2021;40:Article 8.